

The Interpolation Theorem for Classical Logic

Branden Fitelson

09/19/08

Theorem. Let A and B be formulas of P , such that (1) they share at least one propositional symbol in common, and (2) $\models_p A \supset B$. For any two such formulas of P , there exists a formula C (called the P -interpolant of the formulas A and B) such that (3) $\models_p A \supset C$, (4) $\models_p C \supset B$, and (5) C contains only propositional symbols that occur in both A and B (i.e., only propositional symbols shared by A and B).

Setup for an inductive proof. Let $S(A)$ be the set of propositional symbols occurring in A , $S(B)$ be the set of propositional symbols occurring in B , and q be some propositional symbol that is shared by A and B (i.e., $q \in S(A) \cap S(B)$). We will focus on the set $X = S(A) - S(B)$ of propositional symbols that occur in A but not in B . We will prove the interpolation theorem by strong mathematical induction on the cardinality of X . That is, we will prove that the following property of natural numbers holds for all $n \geq 0$:

$\mathfrak{S}(n)$: The interpolation theorem (above) holds when $\overline{X} = \overline{S(A) - S(B)} = n$.

Proof. As always, a proper inductive proof involves a Basis Step and an Inductive step.

Basis Step. Prove $\mathfrak{S}(0)$. That is, we must prove the interpolation theorem for the case in which there are *zero* propositional symbols occurring in A that do not also occur in B ($\overline{X} = 0$). In this case, the set of propositional symbols in A is a subset of those in B [$S(A) \subseteq S(B)$]. Let $C = A$. Then, obviously, (3) $\models_p A \supset C$, since $\models_p A \supset A$. And, since the assumption of the theorem is that $\models_p A \supset B$, we also know that $\models_p C \supset B$. All we need to show is that (5) C contains only propositional symbols that occur in both A and B . But, this follows from the fact that $C = A$, and the assumption of the Basis Step, which is $S(A) \subseteq S(B)$. \square

Inductive Step: Here, we will *assume* as our *inductive hypothesis* that $\mathfrak{S}(m)$ holds when $0 < m < n$. Then, we will *prove from this assumption* that $\mathfrak{S}(n)$ is true. Assume that A and B satisfy conditions (1) and (2) of the theorem, and that there are n propositional symbols $\{p_1, \dots, p_n\}$ occurring in A that do not occur in B (i.e., that $X = \{p_1, \dots, p_n\}$, hence $\overline{X} = n$). Now, define three formulas A_1 , A_2 , and A' , as follows:

A_1 : A with all occurrences of p_n replaced by $q \supset q$.

A_2 : A with all occurrences of p_n replaced by $\sim(q \supset q)$.

A' : $A_1 \vee A_2$. [Note: $P \vee Q \stackrel{\text{def}}{=} \sim P \supset Q$, and $P \wedge Q \stackrel{\text{def}}{=} \sim(P \supset \sim Q)$.]

Next, we use the inductive hypothesis and these three defined formulas to show that there must exist a formula C with the desired properties (3)-(5) required by the theorem. The trick here will be to use the inductive hypothesis on A' and B . We can do this because (i) A' and B share some symbols (at least q) in common, (ii) $\models_p A' \supset B$, and (iii) there are $n - 1 < n$ symbols occurring in A' that do not occur in B . It is obvious that (i) and (iii) are true. However, (ii) requires some argument. We know that $\models_p A \supset B$. It turns out that this allows us to prove $\models_p (A_1 \vee A_2) \supset B$. Remember, the entailment $\models_p A \supset B$ depends only on the *sentential form* of A . And, both A_1 and A_2 *have the same sentential form as* A . As a result, $\models_p A \supset B$ implies $\models_p A_1 \supset B$ and $\models_p A_2 \supset B$, which suffices to establish $\models_p (A_1 \vee A_2) \supset B$. This suffices because $\models_p ((A_1 \supset B) \wedge (A_2 \supset B)) \supset ((A_1 \vee A_2) \supset B)$, which can be verified by truth-table reasoning.

Important Digression on Validity vs Truth-Preservation. The inferences from $A \supset B$ to $A_1 \supset B$ and $A_2 \supset B$ are *validity* preserving, but *not truth*-preserving. All we have shown here is that if $A \supset B$ is true on *all* interpretations, then so is $A_1 \supset B$ (and $A_2 \supset B$). This does *not* imply that every interpretation on which $A \supset B$ is true is also an interpretation on which $A_1 \supset B$ (and $A_2 \supset B$) is true. That is, we have *only* proved

If $\models_p A \supset B$, then $\models_p A_1 \supset B$.

We have *not* proven the following — *nor is the following true in the metatheory of P!*

$$\models_P (A \supset B) \supset (A_1 \supset B).$$

To see that this meta-claim about P is *false*, consider the following counter-example. Let $A = (p'' \supset p''')$, and $B = (p'' \supset p')$. Then, $A \supset B [(p'' \supset p''') \supset (p'' \supset p')]$ is *not* valid [it is F when p' is F, p'' is T, and p''' is T - check this!], but A and B otherwise satisfy the preconditions of the non-trivial case of Craig's theorem [they share one symbol (p'') and there is one symbol (p''') in $S(A) - S(B)$]. While $A \supset B$ is *not valid*, it is true on *some* interpretations. For instance, $A \supset B$ is T whenever p'' is T, and p''' is F (check this!). But, $A_1 \supset B [(p'' \supset (p'' \supset p''')) \supset (p'' \supset p')]$ is F on some of these interpretations. Specifically, $A_1 \supset B$ is F when p'' is T, p''' is F, and p' is F (check this!). So, this shows that the inference from $A \supset B$ to $A_1 \supset B$ is *not truth* preserving, even though it is *validity* preserving. A similar argument can be given to show that the inference from $A \supset B$ to $A_2 \supset B$ is *merely* validity preserving. As I mentioned earlier in the course, all truth preserving inferences are validity preserving. But, as this example explicitly shows, the converse of this entailment in the metatheory of P is *false*. So, there are ways of instantiating A and B such that $\not\models_P (A \supset B) \supset (A_1 \supset B)$. But, these will always be cases in which both $\not\models_P A \supset B$, and $\not\models_P A_1 \supset B$.

Returning to the Inductive Step of our proof, we have just established that (i) A' and B share some symbols (at least q) in common, (ii) $\models_P A' \supset B$, and (iii) there are $n - 1 < n$ symbols occurring in A' that do not occur in B . Thus, the inductive hypothesis applies to A' and B . As such, the inductive hypothesis implies that there must exist a formula C , which *interpolates* A' and B . That is, there must exist a C such that (3') $\models_P A' \supset C$, (4') $\models_P C \supset B$, and (5') C contains only symbols occurring in both A' and B . Moreover, a little more thought reveals that C will *also* be an *interpolant* of A and B ! In order to show that C is also an interpolant of A and B (which will complete the Inductive Step and the proof), we will need to show that C is such that (3) $\models_P A \supset C$, (4) $\models_P C \supset B$, and (5) C contains only symbols occurring in both A and B . Property (4) already follows from the inductive hypothesis, as applied to A' and B [(4')]. So, we just need to establish properties (3) and (5), and we'll be done. Property (5) is easy to establish. We already know (5') that C contains only symbols occurring in both A' and B . And, by construction, $S(A') \subseteq S(A)$. That just leaves property (3) $\models_P A \supset C$. We know that $\models_P A' \supset C$, so if we can prove $\models_P A \supset A'$, then we can infer $\models_P A \supset C$ by a simple *transitivity* argument. Happily, $\models_P A \supset A'$ can be shown by "truth-table reasoning". In the following table, the grayed-out rows are impossible, because of the definitions of A_1 and A_2 (explanations are given). Inspection of the remaining rows reveals that $A \models_P A'$ [$\therefore \models_P A \supset A'$].

A_1	A_2	A	p_n	$A' = A_1 \vee A_2$
T	T	T	T	T
T	T	T	F	T
T	T	F	T	$p_n \models_P A \equiv A_1$
T	T	F	F	$\sim p_n \models_P A \equiv A_2$
T	F	T	T	T
T	F	T	F	$\sim p_n \models_P A \equiv A_2$
T	F	F	T	$p_n \models_P A \equiv A_1$
T	F	F	F	T
F	T	T	T	$p_n \models_P A \equiv A_1$
F	T	T	F	T
F	T	F	T	T
F	T	F	F	$\sim p_n \models_P A \equiv A_2$
F	F	T	T	$p_n \models_P A \equiv A_1$
F	F	T	F	$\sim p_n \models_P A \equiv A_2$
F	F	F	T	F
F	F	F	F	F

That completes the Inductive Step, and with it the inductive proof of the interpolation theorem for P . \square